A Small Sample!

Ames Research Center Activities in AI and Data Science

Nikunj C. Oza, Ph.D.
Leader, Data Sciences Group
Intelligent Systems Division (Code TI)
NASA Ames Research Center
nikunj.c.oza@nasa.gov

The Data Sciences Group at NASA Ames

Machine Learning and Data Mining Research and Development (R&D) for application to NASA problems (Aeronautics, Earth Science, Space Exploration, Space Science)

Group Members

Ilya Avrekh
Kevin Bradner
David Iverson
Miguel Martinho
Bryan Matthews
Milad Memarzadeh, Ph.D.
Nikunj Oza, Ph.D.
Adwait Sahasrabhojanee
Thomas Templin, Ph.D.
Hamed Valizadegan, Ph.D.
Michael von Pohle
Daniel Weckler
+ students

Funding Sources

- NASA Aeronautics Research Mission Directorate- SWS, CAS
- NASA Engineering and Safety Center (NESC)
- Human Research Program (HRP)
- Center Innovation Fund (CIF)
- JPL Advanced Multi-Mission Operations System (AMMOS)

Team Members are NASA Employees, Contractors, and Students.

The Data Sciences Group at NASA Ames

Machine Learning and Data Mining Research and Development (R&D) for application to NASA problems (Aeronautics, Earth Science, Space Exploration, Space Science)

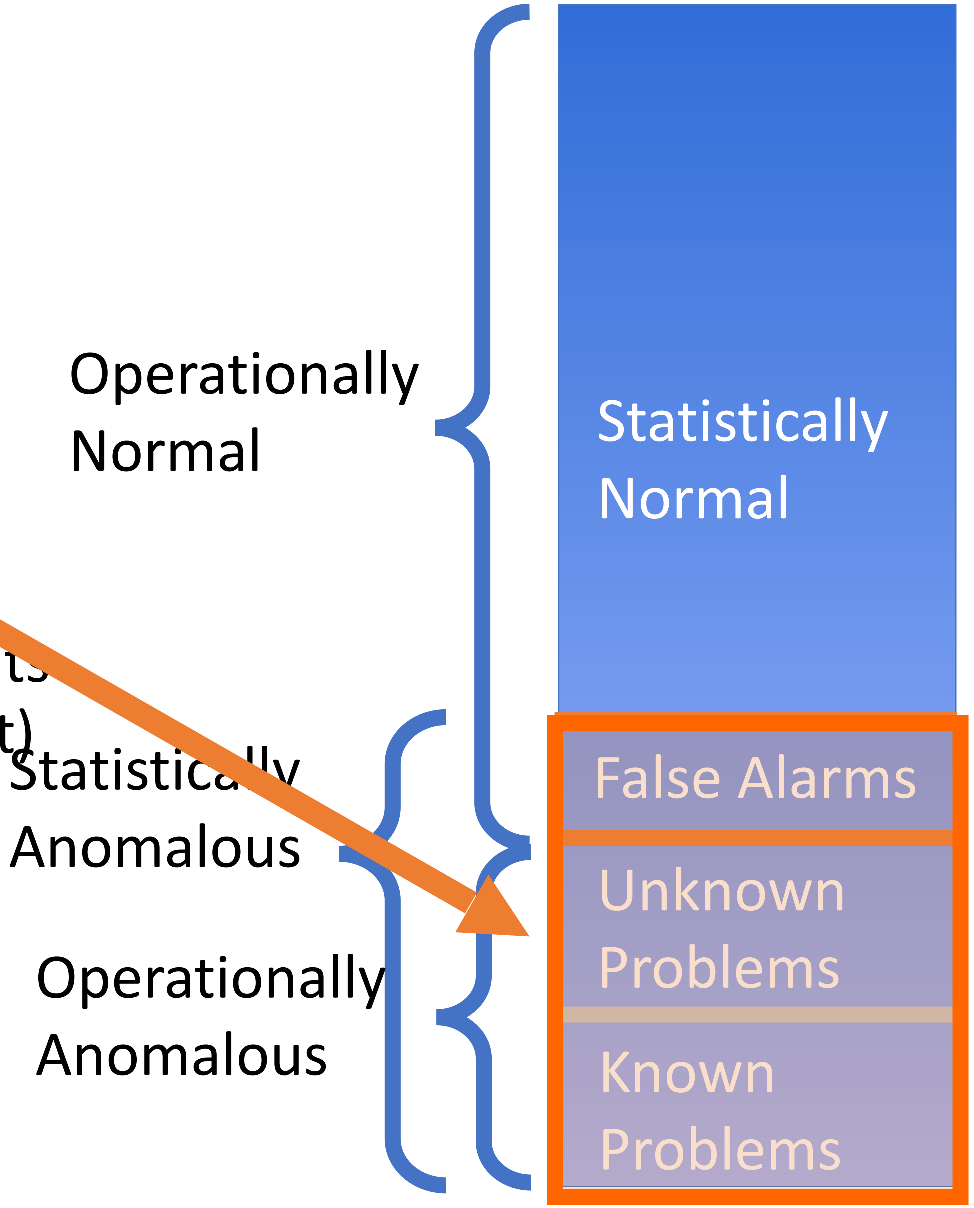
Example Problems / Aeronautics:

- Anomaly Detection
- Precursor Identification
- text mining (classification, topic identification) of safety reports
- relating pilot fatigue to aircraft performance
- identifying patterns in RNAV waypoint compliance
- Surrogate modeling for helicopter noise

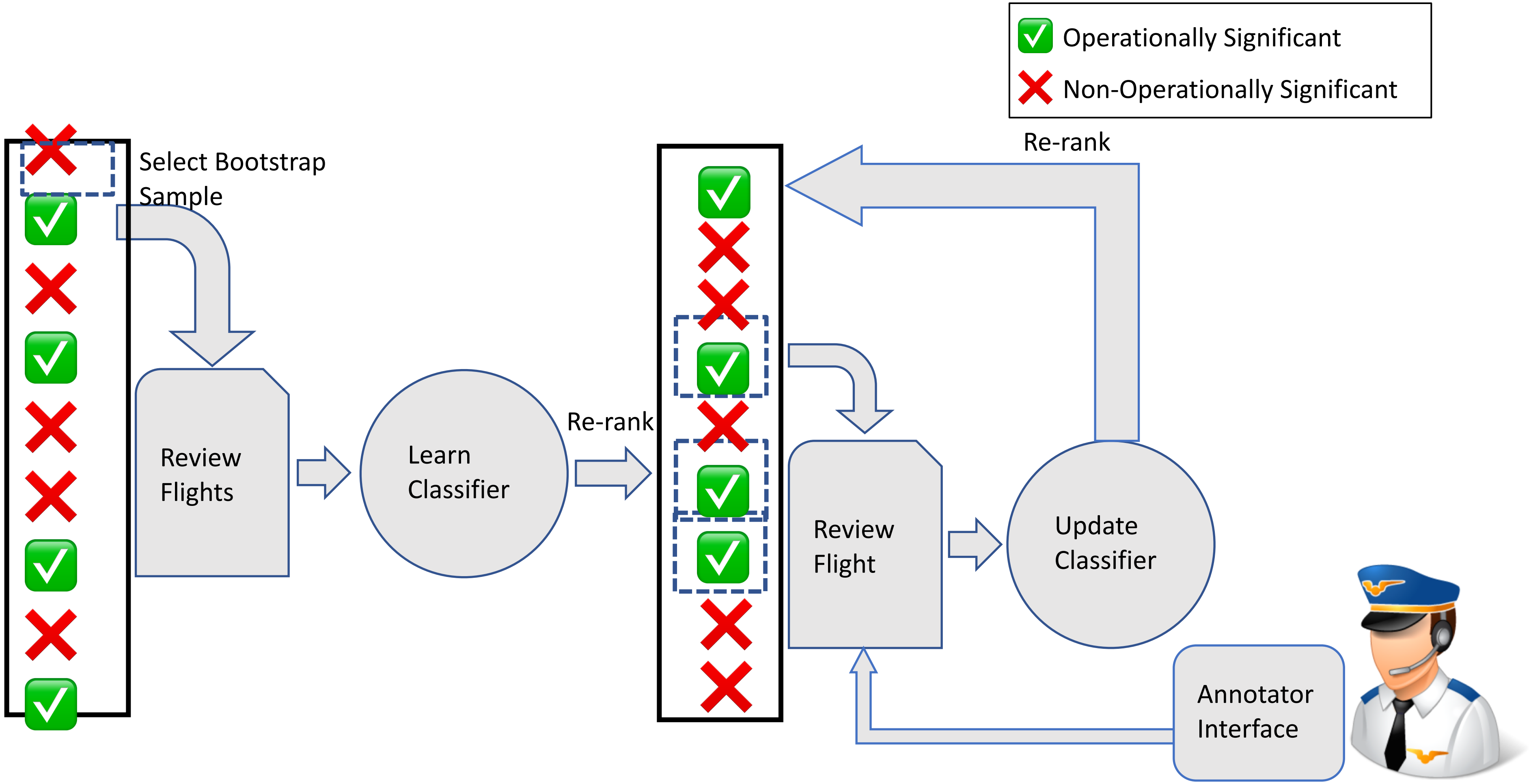
Team Members are NASA Employees, Contractors, and Students.

Data-Driven Methods

- DISCOVER anomalies by
  - learning statistical properties of the data
  - finding which data points do not fit (e.g., far away, low probability)
- Complementary to existing methods
  - Lower false negative (missed detection) rate
  - Higher false positive rate (identified points/flights unusual, but not always operationally significant)
- Data-driven methods -> insights -> modification of exceedance detection



Active Learning Approach



The Data Sciences Group at NASA Ames

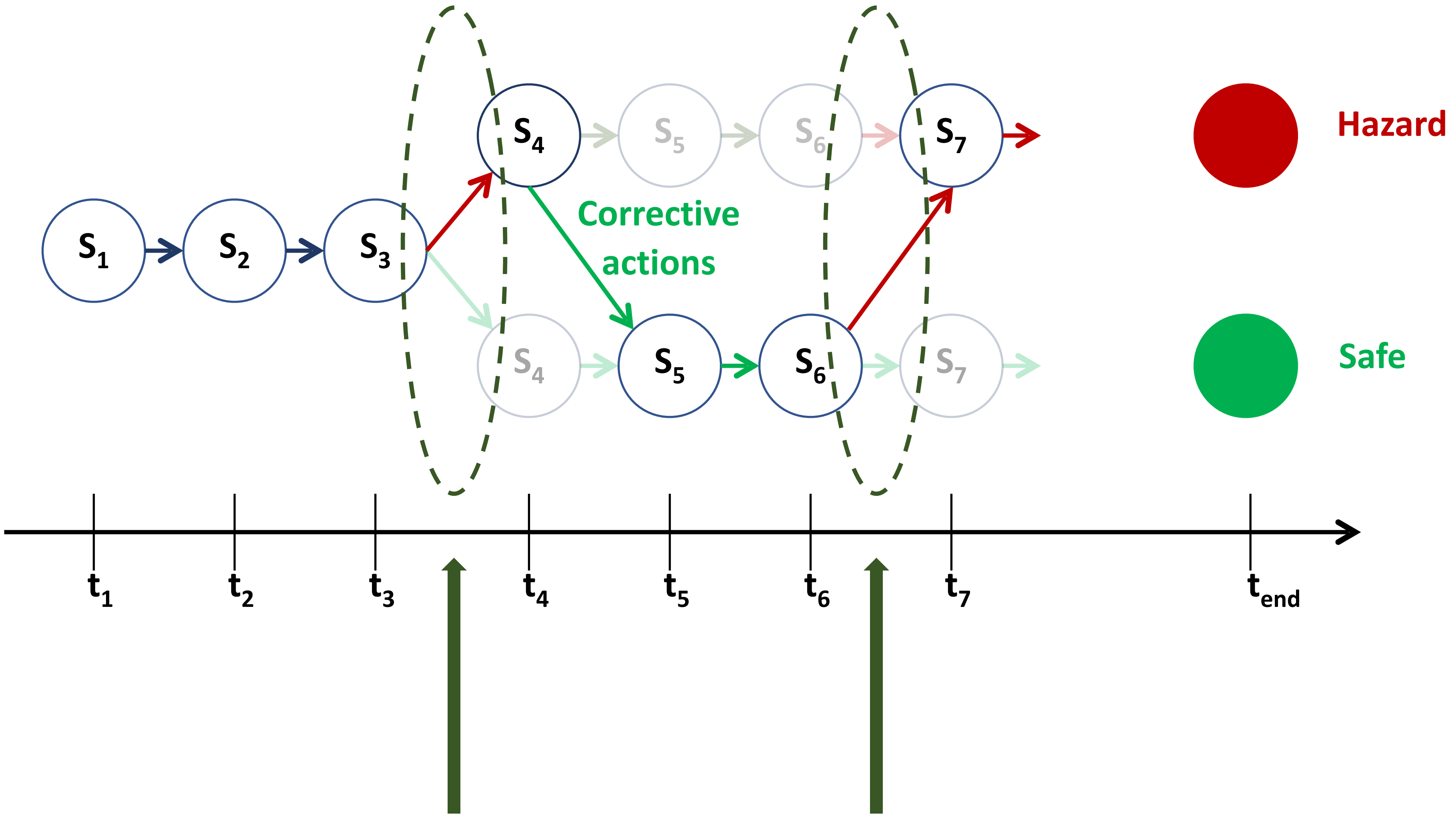
Machine Learning and Data Mining Research and Development (R&D) for application to NASA problems (Aeronautics, Earth Science, Space Exploration, Space Science)

Example Problems / Aeronautics:

- Anomaly Detection
- Precursor Identification
- text mining (classification, topic identification) for commercial aviation
- relating pilot fatigue to aircraft performance
- identifying patterns in RNAV waypoint compliance

Team Members are NASA Employees, Contractors, and Students.

Precursors



S1
S2
S3
S4
S5
S6
S7
S4
S5
S6
S7
Corrective actions
Hazard
Safe
t1
t2
t3
t4
t5
t6
t7
tend

The Data Sciences Group at NASA Ames

Machine Learning and Data Mining Research and Development (R&D) for application to NASA problems (Aeronautics, Earth Science, Space Exploration, Space Science)

Example Problems / Earth Science

- Filling in missing measurements (e.g., ground-based pollution sensors) through relationships with other measurements (e.g., satellite remote sensing)
- anomaly detection
- graph mining to find teleconnections and changes in them
- learning relationships between vegetation and climate variables through symbolic regression

Team Members are NASA Employees, Contractors, and Students.

The Data Sciences Group at NASA Ames

Machine Learning and Data Mining Research and Development (R&D) for application to NASA problems (Aeronautics, Earth Science, Space Exploration, Space Science)

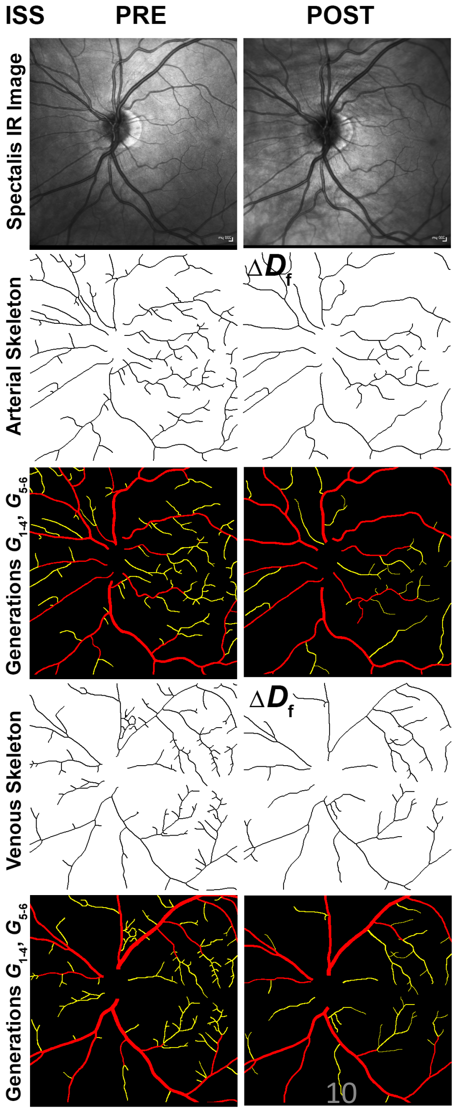
Example Problems / Space

- Space Science: Kepler and TESS planet candidate identification
- Human Space Exploration
  - system health management (monitoring ISS using in-house Inductive Monitoring System)
  - vascular structure identification for astronaut health
  - machine learning within Advanced MultiMission Operations System (AMMOS)

Team Members are NASA Employees, Contractors, and Students.

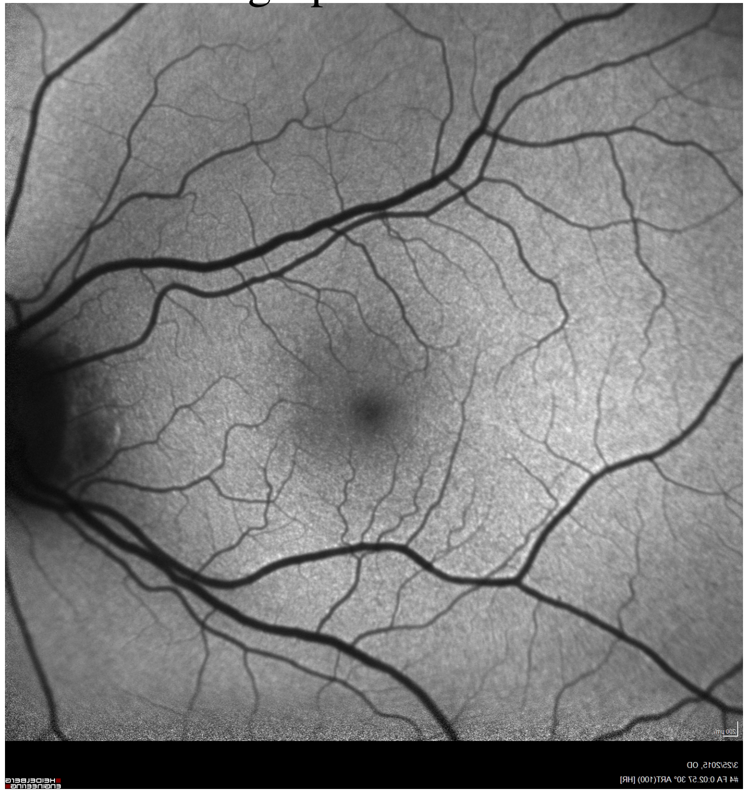
VESGEN (VESsel GENeration) Software

- Developed by Dr. Patricia Parsons and VESGEN lab in Glenn Research Center and later in NASA ARC
- Maps and quantifies vascular morphological characteristics
  - parameters such as diameter, length, branch points, density, and fractal dimension
- Applications
  - Progression of human diabetic retinopathy
  - Remodeling of plant leaf venation patterns in response to plant growth, genetic engineering, and other growth perturbation
  - Progressive Vascular Inflammation in Gastrointestinal Systems (GI): Important for Astronaut Risks in High Radiation Environments
  - Analysis of loss of vessel density in Spaceflight Associated Neuro-ocular Disorder (SANS) (Picture: VESGEN results for ISS Crew Member retina)



Objective

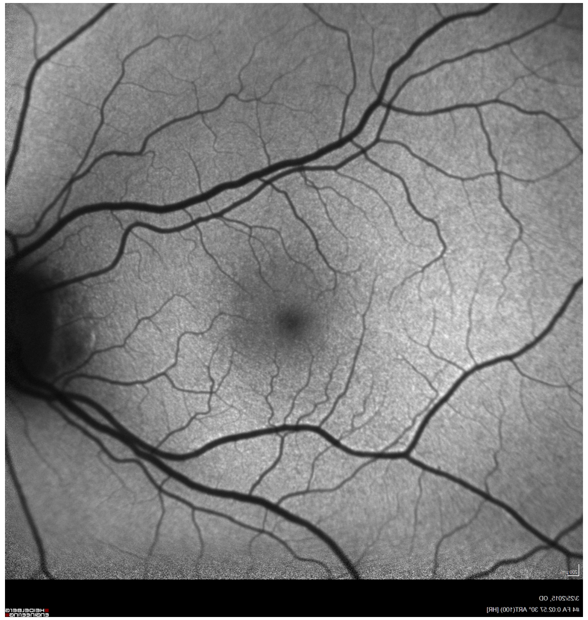
Heidelberg Spectralis IR mode

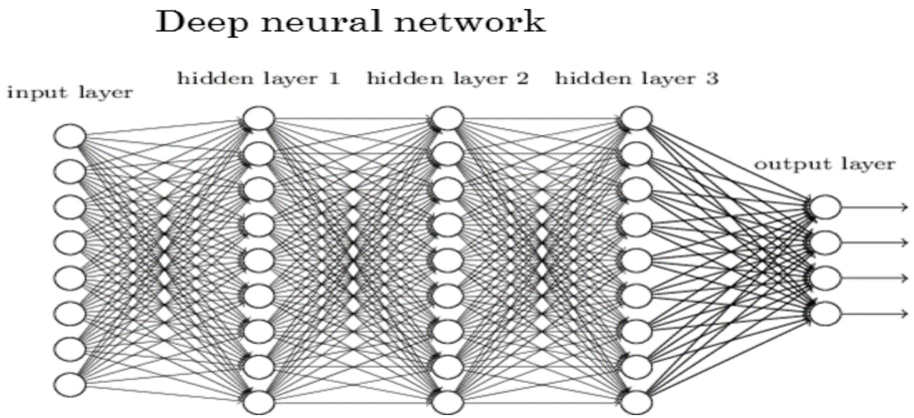


Binarization

- VESGEN requires binary (vascular systems vs. background) image as input
- Currently, manual binarization is being done before feeding the image to VESGEN for analysis
  - Very time consuming and inefficient (2-15 hours image preparation for VESGEN)
- We developed a more automated approach

Supervised Approach: Deep Learning

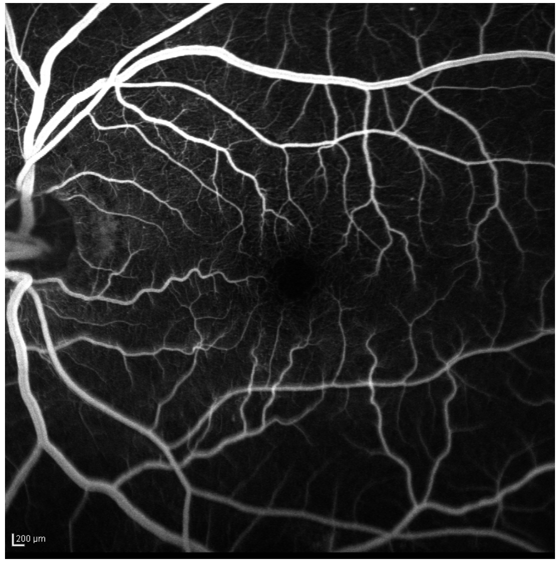




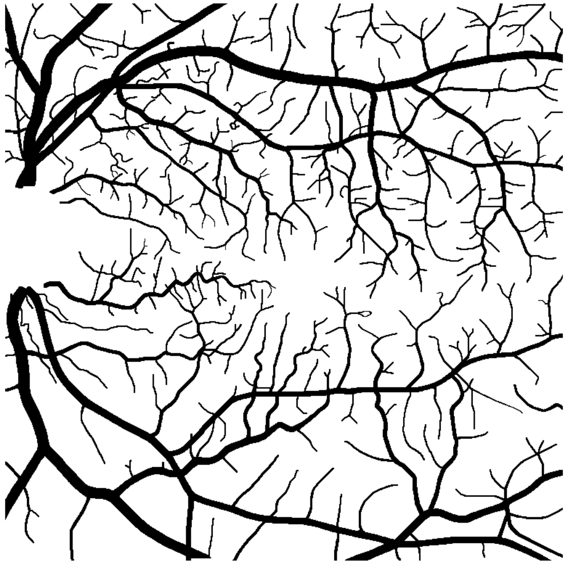
- Assumes existing manually binarized images are provided
- Common practice: whole non-binary image as input; whole binarized (segmented) image as output
  - Learns the mapping from original non-binary image to binary image
  - Such supervised learning models require thousands of images as input/output to learn the task
  - Lack of enough binarized images: 35 vascular images only
- Training time depends on total number of training examples, difficulty of the problem, size of network

Binarization result (NASA data set)

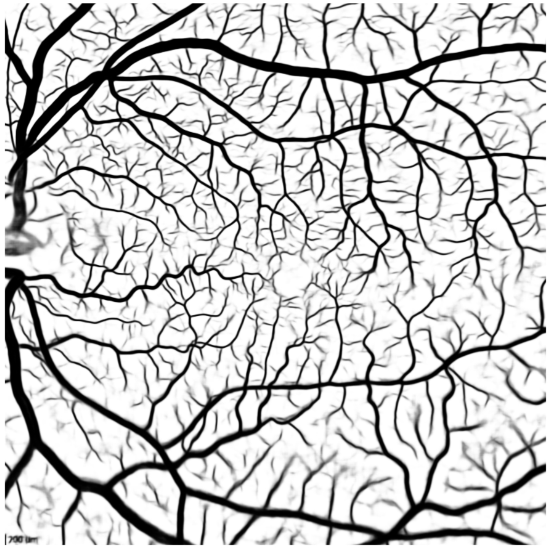
Per Patch	VESGEN
Accuracy (%)	94.61
Precision (%)	81.59
Sensitivity (%)	85.39
Specificity (%)	96.48
AUC	0.9768

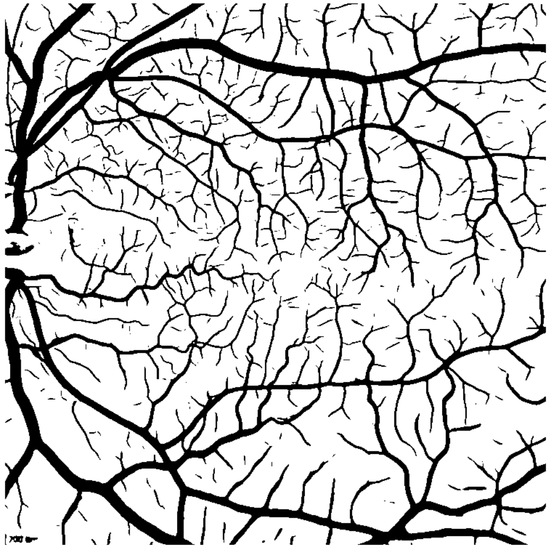


Ground truth

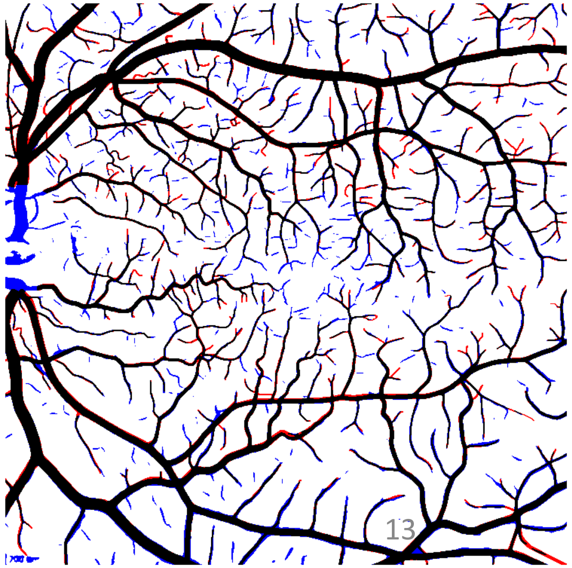


Output heatmap





False positives (blue)/negatives (red)



A Small Sample!

Ames Research Center Activities in AI and Data Science

Nikunj C. Oza, Ph.D.
Leader, Data Sciences Group
Intelligent Systems Division (Code TI)
NASA Ames Research Center
nikunj.c.oza@nasa.gov